



King's Research Portal

DOI:

[10.1109/TASLP.2015.2438547](https://doi.org/10.1109/TASLP.2015.2438547)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

De Sena, E., Hacıhabiboglu, H., Cvetkovic, Z., & Smith, J. (2015). Efficient Synthesis of Room Acoustics via Scattering Delay Networks. *Ieee Transactions On Audio Speech And Language Processing*, 23(9), 1478-1492. <https://doi.org/10.1109/TASLP.2015.2438547>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Efficient Synthesis of Room Acoustics via Scattering Delay Networks

Enzo De Sena, *Member, IEEE*, Hüseyin Hacıhabiboğlu, *Senior Member, IEEE*,
Zoran Cvetković, *Senior Member, IEEE*, Julius O. Smith III, *Member, IEEE*

Abstract—An acoustic reverberator consisting of a network of delay lines connected via scattering junctions is proposed. All parameters of the reverberator are derived from physical properties of the enclosure it simulates. It allows for simulation of unequal and frequency-dependent wall absorption, as well as directional sources and microphones. The reverberator renders the first-order reflections exactly, while making progressively coarser approximations of higher-order reflections. The rate of energy decay is close to that obtained with the image-source method (ISM) and consistent with the predictions of Sabine and Eyring equations. The time evolution of the normalized echo density, which was previously shown to be correlated with the perceived texture of reverberation, is also close to that of ISM. However, its computational complexity is one to two orders of magnitude lower, comparable to the computational complexity of feedback delay network (FDN), and its memory requirements are negligible.

Index Terms—Room acoustics, acoustic simulation, digital waveguide network, reverberation time, echo density.

I. INTRODUCTION

A comprehensive account of the first fifty years of artificial reverberation in [3] identifies three main classes of digital reverberators: delay networks, physical room models and convolution-based algorithms. The earliest class consisted of delay networks, which were the only artificial reverberators feasible with the integrated circuits of the time. The first delay network reverberator, as introduced by Schroeder, was a cascade of several allpass filters, a parallel bank of feedback comb filters and a mixing matrix [3]–[5]. Feedback delay networks (FDNs) were developed as a multichannel extension of the Schroeder reverberator [6], [7], and consist of parallel delay

lines connected recursively through a unitary feedback matrix. The state of the art FDN is due to Jot and Chaigne [8], who proposed using delay lines connected in series with multiband absorptive filters, and a global tonal correction filter to obtain a frequency-dependent reverberation time. FDN reverberators are still among the most commonly used artificial reverberators owing to their extremely low computational complexity and high reverberation quality.

FDNs are structurally equivalent to a particular case of digital waveguide networks (DWNs) [9], reverberators based on the concept of digital waveguides, introduced in [10]. DWNs consist of a closed network of bidirectional delay lines interconnected at lossless junctions. DWNs have an exact physical interpretation as network of interconnected acoustical tubes and have a number of appealing properties in terms of computational efficiency and numerical stability. Reverberators based on delay networks have been widely used for artistic purposes in music production. Artists and sound engineers adjust the available free parameters (e.g., delay line lengths, feedback matrices, absorption coefficients, etc.) until intended qualities of reverberated sound are achieved.

In the domain of predictive architectural modeling, virtual reality and computer games, on the other hand, the objective of artificial reverberators is to emulate the response of a physical room given a set of physically relevant parameters [11], [12]. These parameters include for instance the room geometry, absorption and diffusion characteristics of walls and objects, and position and directivity pattern of source and microphone. We refer to digital reverberators that have this objective as *room acoustics synthesizers* in order to emphasize that the aim is to emulate a specific physical space, rather than achieve certain perceptual qualities, which in general need not have a physical basis.

Various physical models have been proposed in the past for the purpose of room acoustic synthesis. Widely used geometric-acoustic models make the simplifying assumption that sound waves propagate as rays. In particular, the ray tracing approach explicitly tracks the rays emitted by the acoustic source as they bounce off the surfaces of a modeled enclosure. The image-source method (ISM) is an alternative algorithmic implementation that formally replaces the physical boundaries surrounding the source with an equivalent infinite lattice of image sources [13]. Allen and Berkley proved that, in the case of rectangular rooms, this approach is equivalent to solving the wave equation provided that the walls are perfectly rigid [13]. When the walls are not rigid, the results of the ISM

The work reported in this paper was partially funded by (i) EPSRC Grant EP/F001142/1, (ii) European Commission under Grant Agreement no. 316969 within the FP7-PEOPLE Marie Curie Initial Training Network “Dereverberation and Reverberation of Audio, Music, and Speech (DREAMS)”, and (iii) TUBITAK Grant 113E513 “Spatial Audio Reproduction Using Analysis-based Synthesis Methods”.

Enzo De Sena is with ESAT-STADIUS, KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium (e-mail: enzo.desena@esat.kuleuven.be). This work was done in part while he was with CTR, King’s College London.

Zoran Cvetković is with the Centre for Telecommunications Research (CTR), King’s College London, Strand, London, WC2R 2LS, United Kingdom (e-mail: zoran.cvetkovic@kcl.ac.uk).

Hüseyin Hacıhabiboğlu is with Department of Modeling and Simulation, Informatics Institute, Middle East Technical University, Ankara, 06800, Turkey (e-mail: hhuseyin@metu.edu.tr). This work was done in part while he was with King’s College London.

Julius O. Smith III is with CCRMA, Stanford University, Stanford, CA 94304, USA.

Part of this work was presented at the 41st International Conference of the Audio Engineering Society [1], and SITIS 2011 [2].

are no longer physically accurate, but the method still retains its geometric-acoustic interpretation. The main advantage of geometric-acoustic models in comparison to other physical models is their lower – although still considerable – computational complexity. However, they do not model important wave phenomena such as diffraction and interference. These phenomena are inherently modeled by methods based on time and space discretization of the wave equation, such as finite-difference time-domain (FDTD) and digital waveguide mesh (DWM) models [12], [14]–[16]. The main limitation of physical room models consists of their significant computational and memory requirements. While this may not be problematic for predictive acoustic modeling applications, it is a significant limitation for interactive applications, such as virtual reality (VR). Similarly, convolutional methods, which operate by filtering anechoic audio samples with measured room impulse responses (RIRs), do not allow interactive operation unless interpolation among an extensive number of RIRs is supported.

Virtual reality has become a widespread technology, with applications in military training, immersive virtual musical instruments, archaeological acoustics, and, in particular, computer games. Along with realistic graphics rendering, spatial audio is one of the most important factors that affect how realistic its users perceive a virtual environment [17]. The aural and visual components should be consistent and mutually supportive so as to minimise cross-modal sensory conflicts [17].

Room acoustic synthesis for VR also requires flexibility in terms of audio output devices. For example, a full scale VR suite such as a CAVE automatic virtual environment (CAVE) can use Ambisonics [18] or wave field synthesis (WFS), which requires from tens to hundreds of loudspeaker channels [19]. In contrast, a portable game console has a two-channel output that may be used, for instance, to reproduce binaural audio over headphones [20]. Typical home users, on the other hand, commonly use stereophonic, 5.1 or 7.1 setups, whereas the ultra high definition TV (UHDTV) standard, also aimed at home users, makes provisions for 22.2 setups [21].

In summary, room acoustic synthesizers for VR require (i) explicit (tuning-free) modeling of a given virtual space, (ii) scalability in terms of reproduction medium, and (iii) low computational complexity. On the other hand, of the three classes of digital reverberators, (i) convolutional methods do not allow interactive operation without extensive tabulation and interpolation, (ii) delay network methods do not model explicitly a given virtual space and require tuning, and (iii) physical models have a high computational cost.

In order to combine the appealing properties of delay networks and physical models, one possible approach consists of designing delay networks that have parameters with an explicit physical interpretation. Studies in [9], [22] and [23] follow this direction. In [9] and [22] the length of the delay lines of FDNs are chosen such that the lowest eigenfrequencies of the room were reconstructed exactly. In [23], a DWN with few junctions is used to model a rectangular room. The rationale behind the design is to aggressively prune-down a DWM in order to reduce the complexity while retaining an acceptable perceptual result. This approach has various advantages, however the network requires careful manual tuning in order to provide

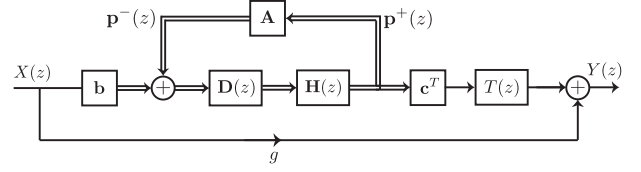


Fig. 1. Block diagram of the modified FDN reverberator as proposed by Jot and Chaigne [8].

satisfactory results [23].

In this work, following the same concept of DWN structures as studied by Karjalainen *et al.* in [23], we propose an architecture that has a number of appealing properties. The proposed architecture, which we refer to as scattering delay network (SDN), renders the line-of-sight component and first-order early reflections accurately both in time and amplitude, while producing a progressively coarser approximation of higher-order reflections. For this reason, SDN can be interpreted as an approximation to geometric acoustics models [13], [24]. SDNs thus approach the accuracy of full-scale room simulation while maintaining computational efficiency on par with typical delay network-based methods. Furthermore, the parameters of SDN are inherited directly from room geometry and absorption properties of wall materials, and therefore do not require ad hoc tuning.

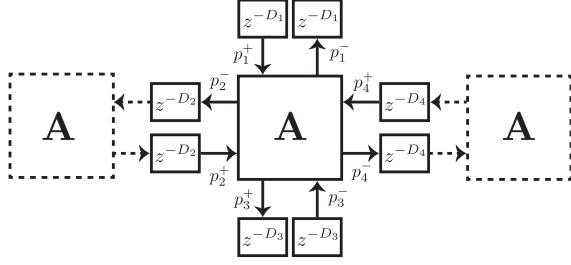
The paper is organized as follows: Section II presents a brief overview of FDNs, DWNs and models proposed by Karjalainen *et al.* [23]. Section III describes the proposed SDN method. The properties of SDNs are studied in Section IV. Section V presents numerical evaluation results. Section VI concludes the paper.

II. BACKGROUND

The proposed SDN reverberator draws inspiration from DWN and DWM structures, however it is in essence a recursive linear time-invariant system. Hence, to provide a comprehensive context, in this section we briefly review FDNs, which are the most commonly used recursive linear time-invariant reverberators, followed by a more detailed review of relevant DWN and DWM material.

A. Feedback Delay Networks

The state-of-the-art feedback delay network (FDN) form, as proposed by Jot and Chaigne [8], is shown in Fig. 1. Here, \mathbf{b} and \mathbf{c} are input and output gains, respectively, $\mathbf{D}(z) = \text{diag}(z^{-m_1}, z^{-m_2}, \dots, z^{-m_N})$ are integer delays, $\mathbf{H}(z) = \text{diag}(H_1(z), \dots, H_N(z))$ are absorption filters, $T(z)$ is the tone correction filter, g is the gain of the direct path, and \mathbf{A} is the feedback matrix. To achieve a high-quality late reverberation, the feedback loop should be lossless, hence typically the feedback matrix \mathbf{A} is designed to be unitary. Each particular choice of the feedback matrix has corresponding implications on subjective or objective qualities of the reverberator [25]; e.g. in the particular case of the identity matrix, $\mathbf{A} = \mathbf{I}$, the FDN structure reduces to N comb filters connected in parallel and acts as the Schroeder reverberator [5]. Note,

Fig. 2. Operation of a DWN around a junction with $K = 4$ waveguides.

however that unitary matrices are only a subset of possible lossless feedback matrices [9]; we elaborate on this point in the next subsection.

B. Digital Waveguide Networks

DWNs consist of a closed network of digital waveguides [10]. A digital waveguide is made up of a pair of delay lines, which implement the digital equivalent of the d'Alembert solution of the wave equation in a one-dimensional medium. The digital waveguides are interconnected at junctions, characterised by corresponding scattering matrices. Fig. 2 shows an example of four digital waveguides with length D_1, \dots, D_4 samples that meet at a junction with scattering matrix \mathbf{A} . In general, a junction scatters incoming wave variables $\mathbf{p}^+ = [p_1^+, \dots, p_K^+]^T$ to produce outgoing wave variables $\mathbf{p}^- = [p_1^-, \dots, p_K^-]^T$ according to $\mathbf{p}^- = \mathbf{A}\mathbf{p}^+$. Note that if all digital waveguides are terminated by an ideal non-inverting reflection, the DWN is structurally equivalent to an FDN with feedback matrix \mathbf{A} and delay-line lengths of $2D_1, \dots, 2D_4$ samples [9].

DWN junctions are lossless. In this context, *losslessness* is defined according to classical network theory [26]. In particular, a junction with scattering matrix \mathbf{A} is said to be lossless if the input and output *total complex power* are equal:

$$\mathbf{p}^{+*} \mathbf{Y} \mathbf{p}^+ = \mathbf{p}^{-*} \mathbf{Y} \mathbf{p}^- \Rightarrow \mathbf{A}^* \mathbf{Y} \mathbf{A} = \mathbf{Y} \quad (1)$$

where \mathbf{Y} is a Hermitian positive-definite matrix [9] and $(\cdot)^*$ denotes the conjugate transpose. The quantity $\mathbf{p}^{\pm*} \mathbf{Y} \mathbf{p}^{\pm}$ is the square of the *elliptic norm* of \mathbf{p}^{\pm} induced by \mathbf{Y} . It can be shown that a matrix \mathbf{A} is lossless if and only if its eigenvalues lie on the unit circle and it admits a basis of linearly independent vectors [9]. A consequence of this result is that lossless feedback matrices can be fully parametrized as $\mathbf{A} = \mathbf{T}^{-1} \mathbf{\Lambda} \mathbf{T}$, where \mathbf{T} is any invertible matrix and $\mathbf{\Lambda}$ is any unit-modulus diagonal matrix [9].

The DWN can also be interpreted as a physical model for a network of acoustic tubes. In this case \mathbf{A} assumes a particular form. If we denote by y_i the characteristic admittance of the i -th tube and by v_i the volume velocity of the i -th tube at the junction, the continuity of pressure and conservation of velocity at the junction give [26]

$$p_1 = p_2 = \dots = p_K = p, \quad (2)$$

$$v_1 + v_2 + \dots + v_K = 0, \quad (3)$$

respectively, where $p_i = p_i^+ + p_i^-$ denotes the acoustic pressure of the i -th tube. Equations (2) and (3) imply that the pressure p at the junction is given by

$$p = \frac{2}{\sum_{i=1}^K y_i} \sum_{i=1}^K y_i p_i^+ = \frac{2}{\sum_{i=1}^K y_i} \sum_{i=1}^K y_i p_i^-, \quad (4)$$

where we used $v_i = v_i^+ + v_i^-$ and Ohm's law for traveling waves $v_i^+ = y_i p_i^+$ and $v_i^- = -y_i p_i^-$ [26]. Since $p_i^- = p - p_i^+$, the scattering matrix can be expressed as

$$\mathbf{A} = \frac{2}{\langle \mathbf{1}, \mathbf{y} \rangle} \mathbf{1} \mathbf{y}^T - \mathbf{I}, \quad (5)$$

where $\mathbf{1} = [1, \dots, 1]^T$, $\mathbf{y} = [y_1, \dots, y_K]^T$, $\langle \cdot, \cdot \rangle$ denotes the scalar product, and \mathbf{I} is the identity matrix. Observe that the scattering matrix in (5) satisfies equation (1) with $\mathbf{Y} = \text{diag}\{y_1, \dots, y_K\}$ and is therefore lossless. In this physically rooted case, the square of the elliptic norm of \mathbf{p}^{\pm} induced by \mathbf{Y} has the meaning of incoming/outgoing acoustic power: $\mathbf{p}^{\pm*} \mathbf{Y} \mathbf{p}^{\pm} = \sum_{i=1}^K y_i |p_i^{\pm}|^2 = \pm \sum_{i=1}^K v_i^* p_i^{\pm}$ [9].

An equivalent formulation of DWNs involves normalized pressure waves, defined as $\tilde{p}_i^{\pm} = p_i^{\pm} \sqrt{y_i}$. In this case, the propagating wave variables \tilde{p}_i^{\pm} represent the square root of the traveling signal power [9]. If we define $\tilde{\mathbf{Y}} = \text{diag}(\sqrt{y_1}, \dots, \sqrt{y_K})$, the normalized output wave can be written as $\tilde{\mathbf{p}}^- = \tilde{\mathbf{Y}} \mathbf{p}^- = \tilde{\mathbf{Y}} \mathbf{A} \mathbf{p}^+ = \tilde{\mathbf{Y}} \mathbf{A} \tilde{\mathbf{Y}}^{-1} \tilde{\mathbf{p}}^+ = \tilde{\mathbf{A}} \tilde{\mathbf{p}}^+$, where $\tilde{\mathbf{A}} = \tilde{\mathbf{Y}} \mathbf{A} \tilde{\mathbf{Y}}^{-1}$. The equivalent scattering matrix $\tilde{\mathbf{A}}$ can be expressed as

$$\tilde{\mathbf{A}} = \frac{2}{\|\tilde{\mathbf{y}}\|^2} \tilde{\mathbf{y}} \tilde{\mathbf{y}}^T - \mathbf{I}, \quad (6)$$

which is a Householder reflection around the vector $\tilde{\mathbf{y}}$. Such Householder matrices will be also used in the context of SDN reverberators, proposed in the next section, where they will exhibit some sought after properties, including low computational complexity and desirable normalised echo density profiles.

In order to inject energy in a DWN, various methods have been used in the past, ranging from attaching an additional waveguide where the outgoing wave is ignored [10] or using an adapted impedance such that there is no energy reflected along the outgoing wave, to more complex approaches [27]. A common approach is to apply an external ideal volume velocity source to the junction [16]. This is equivalent to superimposing source pressure, p_S , to the pressure due to the waveguides meeting at the junction, p , thus making the total pressure at the junction equal to:

$$\bar{p} = p_S + p. \quad (7)$$

In the context of FDTD models, a source that injects energy in this way is called a *soft-source*, as opposed to a *hard-source*, which actively interferes with the propagating pressure field [28], [29]. In order to implement equation (7) in the DWN structure, the input from the source can be distributed uniformly to incoming wave variables according to

$$\bar{\mathbf{p}}^+ = \mathbf{p}^+ + \frac{p_S}{2} \mathbf{1}. \quad (8)$$

which provides the intended node pressure [30]. This is the approach that will be used for injecting source energy in the proposed SDN structures.

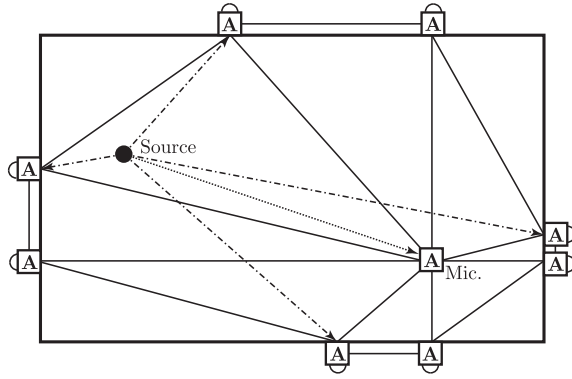


Fig. 3. Conceptual depiction of one of the DWN topologies proposed by Karjalainen *et al.* as seen by an observer above the simulated enclosure (in this case a 2D rectangular room) [23]. The solid black lines denote the bidirectional delay lines interconnecting wall nodes. The dash-dotted lines denote the unidirectional absorptive delay lines. The dotted line denotes the line-of-sight (LOS) component. The solid arcs around junctions denote loaded self-connections implementing losses.

C. Digital Waveguide Meshes

DWNs formed of fine grids of scattering junctions, referred to as *digital waveguide meshes*, are used to model wave propagation in an acoustic medium [31]. Each spatial sample in a digital waveguide mesh (DWM) is represented by a K -port scattering junction, connected to its geometric neighbors over bidirectional unit-sample delay lines. In the typical case of an isotropic medium, \mathbf{y} is a constant vector, while \mathbf{A} and $\tilde{\mathbf{A}}$ are identical and given by

$$\mathbf{A} = \frac{2}{K} \mathbf{1}\mathbf{1}^T - \mathbf{I}. \quad (9)$$

We will refer to such a scattering matrix as the *isotropic scattering matrix*.

In the one-dimensional band-limited case, the DWM model provides the exact solution of the wave equation [26]. In the two [32] and three [33] dimensional cases, sound propagates at slightly different speeds in different directions and different frequencies, causing a *dispersion error* [34], which can be controlled and reduced to some extent by means of careful design of the mesh topology or by using interpolated meshes and frequency warping methods [15], [35].

Accurate modeling with DWMs requires mesh topologies with a very fine resolution (e.g. $\approx 10^7$ junctions for a room of size $4 \times 6 \times 3$ m [23]). That makes the computational load and the amount of memory required prohibitively high for real-time operation, especially for large rooms. These drawbacks motivated the work of Karjalainen *et al.* in [23], which is reviewed in the next subsection.

D. Reduced Digital Waveguide Meshes

In order to lower the computational complexity of DWM models, Karjalainen *et al.* considered coarse approximations of room response synthesis via sparse DWM structures [23]. One such structure is shown in Fig. 3. In this network the sound source is connected via unidirectional absorbing delay lines to scattering junctions. These junctions are positioned

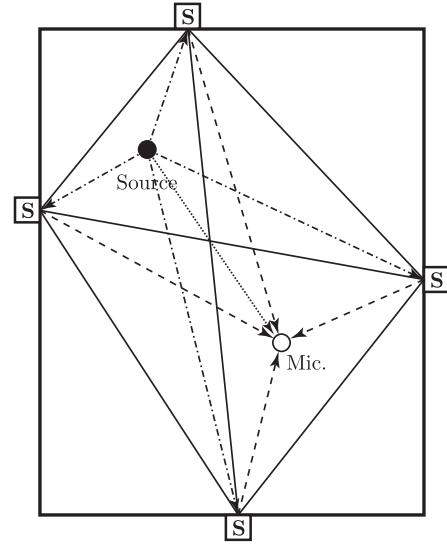


Fig. 4. Conceptual depiction of the SDN reverberator. The solid black lines denote bidirectional delay lines interconnecting the SDN wall nodes. The dash-dotted lines denote unidirectional absorptive delay lines connecting the source to the SDN nodes. The dashed lines denote unidirectional absorptive delay lines connecting the SDN nodes to the microphone. The dotted line denotes the LOS component.

at the locations where first-order reflections impinge on the walls. This ensures that delays of first-order reflections are rendered accurately. The junctions are connected via bidirectional delay lines with the microphone, which is also modeled as a scattering junction contributing to the energy circulation in the network. The line-of-sight component is modeled by a direct connection between the source and the microphone. Additional bidirectional delay lines parallel to the wall edges are included to better simulate room axial modes [36]. All the wall junctions are connected in a ring topology.

In order to model losses at the walls, it appears that a combination of junction loads and additional self-connections is used. However, implementation details are not given, and the authors state that the network required careful heuristic tuning [23]. For these reasons the results are difficult to replicate.

In the next section, we describe a structure which renders the direct path and first-order early reflections accurately both in time and amplitude, while producing progressively coarser approximations of higher order reflections and late reverberation. The proposed method has parameters that are inherited directly from room geometry and absorption properties of wall materials, and thus does not require tuning. In order to distinguish the proposed structure from the ones considered by Karjalainen *et al.* in [23], and because of the importance of the scattering operation, we refer to the proposed structures as *scattering delay networks* (SDNs).

III. SCATTERING DELAY NETWORKS

The aim of the SDN structure is to simulate the acoustics of an enclosure using a minimal topology which would ensure that each significant reflection in the given space has a corresponding reflection in the synthesized response. This requires

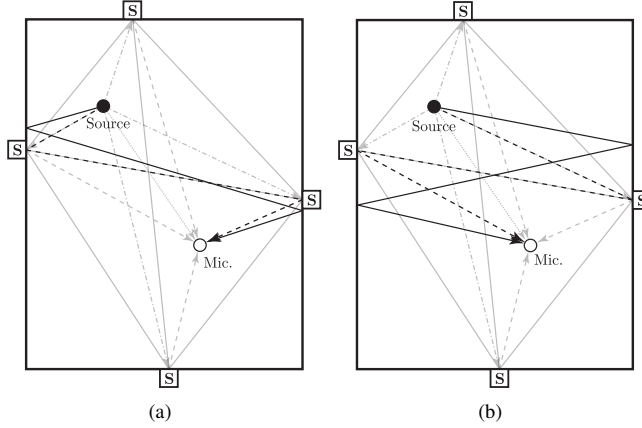


Fig. 5. Examples of approximations generated by SDN for two second-order reflections. The solid black line represents the actual path of the second-order reflection, while the dashed line is the corresponding path within the SDN.

representing each significant reflective surface using one node in a fully connected mesh topology. The concept is illustrated in Fig. 4. For clarity, considerations in this paper will pertain to rectangular empty rooms, however all the presented concepts can be extended in a straightforward manner to arbitrary polyhedral spaces with additional reflective surfaces in their interior.

As shown in Fig. 4, the network consists of a fully connected DWN with one scattering node for each wall. This network is minimal in the sense that the removal of any of the nodes or paths would make it impossible to model a significant subset of reflections which would arise in the space. The source is connected to the scattering nodes via unidirectional absorbing lines. As opposed to the reverberators proposed in [23], the microphone node is a passive element that does not participate in the energy recirculation, hence scattering nodes are connected to the microphone via unidirectional absorbing lines and no energy is injected from the microphone node back to the network.

Early reflections are known to strongly contribute to one's perception of the size and spaciousness of a room [37]. For this reason, nodes are positioned on the walls at locations of first-order reflections. Delays and attenuation of the lines connecting the nodes, source, and microphone are set so that first-order reflections are rendered accurately in their timing and energy.

Second-order reflections are approximated by corresponding paths within the network. This is illustrated in Fig. 5. It can be observed from the figure that the accuracy of second-order reflections depends on the particular reflection, but nonetheless the delays of the approximating paths are similar to the actual ones. As the reflection order increases, coarser approximations are made. The proposed network thus provides a gracefully degrading approximation of geometric-acoustic methods such as ISM and ray-tracing.

Precise details of the SDN design are given below.

Scattering nodes: Each node is positioned on a wall of the modeled enclosure. The nodes are positioned at the locations where the first-order reflections impinge on the walls.

These locations are straightforward to calculate for simple geometries, e.g. convex polyhedra. The nodes carry out a scattering operation on the inputs from the other K nodes, \mathbf{p}^+ , to obtain the outputs as $\mathbf{p}^-[n] = \mathbf{S}\mathbf{p}^+[n]$, where \mathbf{S} is the $K \times K$ (not-necessarily lossless) scattering matrix. Cuboid rooms, which are used in the following, correspond to $K = 5$. Other geometries where K is a power of 2 are computationally convenient since it can be shown that they lead to a multiplier-free realization [10].

The scattering matrix \mathbf{S} governs how energy circulates within the network. Since each node is associated with a wall, it describes how energy is exchanged between walls. Furthermore, when incident waves \mathbf{p}^+ are scattered from the wall, a certain amount of energy is absorbed. A macroscopic quantity that describes material absorption sufficiently well for most synthesis applications is the random-incidence absorption coefficient [38]. This coefficient, specified by the ISO 354 standard, is known for a variety of materials [38].

The wall absorption effect can be expressed in the most general form as $\mathbf{p}^- \mathbf{Y} \mathbf{p}^- = (1 - \alpha) \mathbf{p}^+ \mathbf{Y} \mathbf{p}^+$, where α is the wall absorption coefficient, which is equivalent to

$$\mathbf{S}^* \mathbf{Y} \mathbf{S} = (1 - \alpha) \mathbf{Y}. \quad (10)$$

By expressing \mathbf{S} as $\mathbf{S} = \beta \mathbf{A}$, where $\beta = \sqrt{1 - \alpha}$, the relationship in (10) becomes equivalent to $\mathbf{A}^* \mathbf{Y} \mathbf{A} = \mathbf{Y}$, i.e. the scattering matrix \mathbf{S} is the product of the wall reflection coefficient β and a lossless scattering matrix \mathbf{A} . As in DWNs, if \mathbf{A} is selected as given by (5) or (6) the propagating variables have a physical interpretation as the pressure or root-power waves in a network of acoustic tubes. Other non-physical choices of \mathbf{A} are possible, as long as the lossless condition is satisfied; these are discussed in Section IV-C.

As will be shown in Section V-A, setting the scattering matrix as $\mathbf{S} = \beta \mathbf{A}$ results in a rate of energy decay that is consistent with the well-known Sabine and Eyring formulas as well as with the results of the ISM. This can be attributed to the fact that both the average time between successive reflections (i.e. the *mean free path*) and the energy loss at the walls in the virtual SDN network are close to the ones observed in the corresponding physical room.

The absorption characteristic of most real materials is frequency-dependent. This can be modelled by using a more general scattering matrix $\mathbf{S}(z) = \mathbf{H}(z) \mathbf{A}$, where $\mathbf{H}(z) = \text{diag}\{H(z), \dots, H(z)\}$, and $H(z)$ is a wall filter. The absorption coefficients in consecutive octave bands, for a range of materials, are reported in [38]. Standard filter design techniques can be used to fit the frequency response $H(e^{j\omega})$ to these tabulated values. Minimum-phase IIR filters are particularly convenient in this context as they reduce computational load without significantly affecting the phase of simulated reflections [39].

As in conventional DWMs, the pressure at the SDN node is a function of the incoming wave variables, $p_i^+[n]$, from neighboring nodes and the pressure, $p_S[n]$, injected by the source. That is modelled according to (7), and it is illustrated in Fig. 6. Other details of source-to-node connections are discussed next.

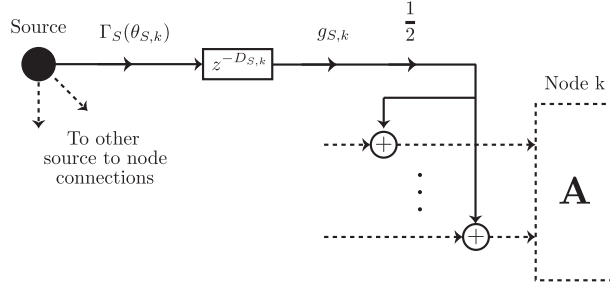


Fig. 6. Connection between the source node and an SDN node.

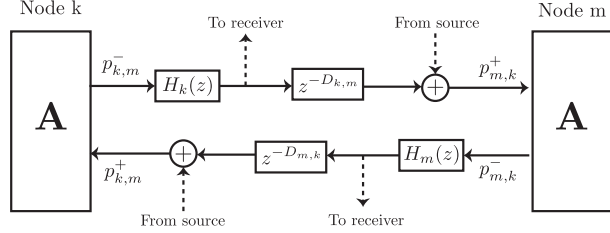


Fig. 7. Two interconnected SDN nodes.

Source-to-node connection: The input to the system is provided by a source node connected to SDN nodes via unidirectional absorbing delay lines (see Fig. 6).

The delay of the line connecting the source at \mathbf{x}_S and the SDN node positioned at \mathbf{x}_k is given by the corresponding propagation delay $D_{S,k} = \lfloor F_s \|\mathbf{x}_S - \mathbf{x}_k\|/c \rfloor$. The attenuation due to spherical spreading ($1/r$ law) is modeled as

$$g_{S,k} = \frac{1}{\|\mathbf{x}_S - \mathbf{x}_k\|}. \quad (11)$$

Source directivity is another important simulation parameter in room acoustic synthesis. The sparse sampling of the simulated enclosure prohibits the simulation of source directivity in detail. However, a coarse approximation can be incorporated by weighting the outgoing signals by $\Gamma_S(\theta_{S,k})$, where $\Gamma_S(\theta)$ is the source directivity, and $\theta_{S,k}$ is the angle between the source reference axis and the line connecting the source and k -th node. An alternative approach consists of using an average of the directivity pattern in some angular sector.

Node-to-node connection: The connections between the SDN nodes consist of bidirectional delay lines modeling the propagation path delay as shown in Fig. 7. Additional low-pass filters can be inserted into the network at this point to model the frequency-dependent characteristic of air absorption, as studied by Moorer [40].

The delays of the lines connecting the nodes are determined by their spatial coordinates. Thus, the delay of the line between a node at location \mathbf{x}_k and a node at \mathbf{x}_m is $D_{k,m} = \lfloor F_s \|\mathbf{x}_k - \mathbf{x}_m\|/c \rfloor$, where c is the speed of sound and F_s is the sampling frequency.

Node-to-microphone connection: Each node is connected to the microphone node via a unidirectional absorbing delay line. The signal extracted from the junction, $p_e[n]$, is a linear combination of outgoing wave variables, $p_e[n] = \mathbf{w}^T \mathbf{p}^-[n]$,

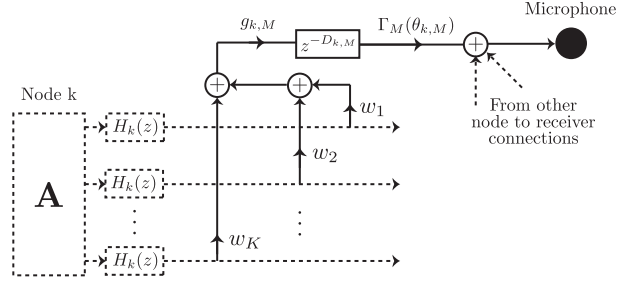


Fig. 8. Connection between an SDN node and the receiver/microphone node.

where $\mathbf{p}^-[n]$ is the wave vector after the wall filtering operation, as shown in Fig. 8.

In the physical case where \mathbf{A} is in the form (5) or (6), the outgoing signal to the microphone is taken as the node's pressure, as given by equations (4) and (7):

$$p_e[n] = \frac{2}{\langle \mathbf{1}, \mathbf{y} \rangle} \mathbf{y}^T \mathbf{p}^-[n]. \quad (12)$$

In the non-physical case, various choices are available for extracting a signal from the junction. The only condition that the weights \mathbf{w} need to satisfy is that the cascade of pressure injection, scattering and extraction does not alter the amplitude of first-order reflections. Since the incoming wave vector \mathbf{p}^+ for a first-order reflection with amplitude p_S is given by $\mathbf{p}^+ = (\frac{p_S}{2} \mathbf{1})$ (see Fig. 6) and since $\mathbf{p}^- = \mathbf{S} \mathbf{p}^+$, this condition can be written as $\mathbf{w}^T \mathbf{S} (\frac{p_S}{2} \mathbf{1}) = \beta p_S$ or, equivalently, as

$$\mathbf{w}^T \mathbf{A} \mathbf{1} = 2. \quad (13)$$

A possible choice for \mathbf{w} is a constant vector, which is computationally convenient since it requires a single multiplication. In this case, the constraint (13) yields the unique solution $\mathbf{w} = \frac{2}{\mathbf{1}^T \mathbf{A} \mathbf{1}} \mathbf{1}$.

The delay from the k -th SDN node to the microphone node is $D_{k,M} = \lfloor F_s \|\mathbf{x}_k - \mathbf{x}_M\|/c \rfloor$. As with the source directivity, the microphone directivity is also modeled using a simple gain element $\Gamma_M(\theta_{k,M})$, where $\Gamma_M(\theta)$ is the microphone directivity pattern and $\theta_{k,M}$ is the angle between the microphone acoustical axis and the k -th node. This approach ensures that the microphone is emulated correctly for the directions associated to the first-order reflections. As with the source-associated connection, an alternative approach consists of using an average of the directivity pattern in some angular sector.

The attenuation coefficient is set as

$$g_{k,M} = \frac{1}{1 + \frac{\|\mathbf{x}_k - \mathbf{x}_M\|}{\|\mathbf{x}_S - \mathbf{x}_k\|}}, \quad (14)$$

such that, using (11),

$$g_{S,k} \times g_{k,M} = \frac{1}{\|\mathbf{x}_S - \mathbf{x}_k\| + \|\mathbf{x}_k - \mathbf{x}_M\|}, \quad (15)$$

which yields the correct attenuation for the first-order reflection according to the $1/r$ law. Notice that the above choice of $g_{k,M}$ and $g_{S,k}$ is not unique in satisfying the constraint (15). The attenuation can in fact be distributed differently between the source-to-node and node-to-microphone branches but with little impact on overall energy decay.

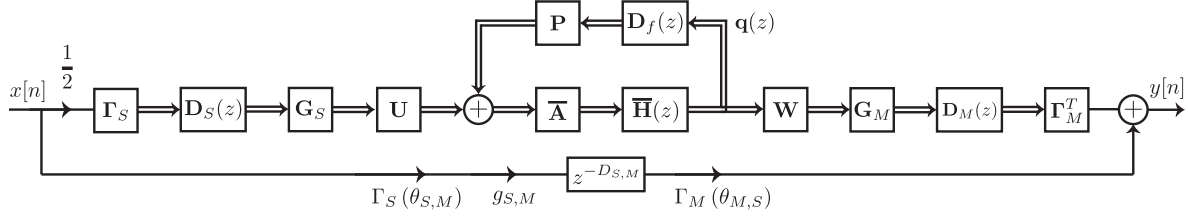


Fig. 9. Block diagram of the SDN reverberator.

IV. PROPERTIES OF SDNs

A. Transfer function

The block diagram of an SDN system is shown in Fig. 9. In the figure,

$$\Gamma_S = [\Gamma_S(\theta_{S,1}), \Gamma_S(\theta_{S,2}), \dots, \Gamma_S(\theta_{S,K+1})]^T, \quad (16)$$

$$\mathbf{D}_S(z) = \text{diag} \{z^{-D_{S,1}}, z^{-D_{S,2}}, \dots, z^{-D_{S,K+1}}\}, \quad (17)$$

$$\mathbf{G}_S = \text{diag} \{g_{S,1}, g_{S,2}, \dots, g_{S,K+1}\}, \quad (18)$$

are the source directivity vector, the source delay matrix, and the source attenuation matrix, respectively. The corresponding quantities associated with the microphone Γ_M , $\mathbf{D}_M(z)$ and \mathbf{G}_M are defined as in (16), (17) and (18), by substituting S with M . Further,

$$\mathbf{U} = \text{diag} \left\{ \underbrace{1, \dots, 1}_{K+1} \right\}, \quad (19)$$

$$\mathbf{D}_f(z) = \text{diag} \{z^{-D_{f,1}}, \dots, z^{-D_{f,K+1}}\}, \quad (20)$$

$$\bar{\mathbf{H}}(z) = \text{diag} \left\{ \underbrace{H_1(z), \dots, H_1(z)}_K, \dots, H_{K+1}(z) \right\}, \quad (21)$$

are the block-diagonal matrix that distributes source signals to input wave variables, the inter-node delay matrix, and the wall absorption matrix, respectively. The block diagonal scattering matrix $\bar{\mathbf{A}}$ and the block diagonal weight matrix \mathbf{W} are defined as

$$\bar{\mathbf{A}} = \text{diag} \{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{K+1}\}, \quad (22)$$

$$\mathbf{W} = \text{diag} \{\mathbf{w}_1^T, \dots, \mathbf{w}_{K+1}^T\}, \quad (23)$$

where \mathbf{A}_k and \mathbf{w}_k are the i -th node's scattering matrix and pressure extraction weights, respectively. While these variables can in general be different for each node, in all the simulations presented in this paper they are selected to be equal, $\mathbf{A}_1 = \dots = \mathbf{A}_{K+1}$ and $\mathbf{w}_1 = \dots = \mathbf{w}_{K+1}$.

Finally, in Fig. 9, $g_{S,M}$ and $z^{-D_{S,M}}$ are the line-of-sight attenuation and delay, respectively, and \mathbf{P} is a permutation matrix whose elements are determined based on the network topology. Due to the underlying node connectivity being bidirectional, this permutation matrix is symmetric. For the simplest case of a three-dimensional enclosure with a rectangular shape, the permutation matrix takes the form $\mathbf{P} = \delta_{i,f(j)}$, where $\delta_{i,j}$ is the Kronecker delta,

$$f(i) = ((6i - ((i-1))_{N-1}))_{N(N-1)} + 1, \quad (24)$$

and $((\cdot))_N$ is the modulo- N operation. Inspection of Fig. 9 reveals that the system output can be expressed as

$$Y(z) = \Gamma_M^T \mathbf{D}_M(z) \mathbf{G}_M \mathbf{W} \mathbf{q}(z) + \bar{g} z^{-D_{S,M}} X(z) \quad (25)$$

where $\bar{g} = g_{S,M} \Gamma_S(\theta_{S,M}) \Gamma_M(\theta_{M,S})$, and $\mathbf{q}(z)$ is the state vector, given by

$$\mathbf{q}(z) = \frac{1}{2} [\mathbf{I} - \bar{\mathbf{H}}(z) \bar{\mathbf{A}} \mathbf{P} \mathbf{D}_f(z)]^{-1} \bar{\mathbf{H}}(z) \bar{\mathbf{A}} \mathbf{U} \mathbf{G}_S \mathbf{D}_S(z) \Gamma_S X(z). \quad (26)$$

The transfer function can therefore be expressed as

$$H(z) = \frac{1}{K} \mathbf{k}_M^T(z) \left[\bar{\mathbf{H}}^{-1}(z) \bar{\mathbf{A}}^T - \mathbf{P} \mathbf{D}_f(z) \right]^{-1} \mathbf{k}_S(z) + \bar{g} z^{-D_{S,M}}, \quad (27)$$

where $\mathbf{k}_M^T(z) = \Gamma_M^T \mathbf{D}_M(z) \mathbf{G}_M \mathbf{W}$ and $\mathbf{k}_S(z) = \mathbf{U} \mathbf{G}_S \mathbf{D}_S(z) \Gamma_S$.

It may be observed that, unlike FDN reverberators, relevant acoustical aspects, such as the direct path and reflection delays, are modeled explicitly, allowing complete control of source and microphone positions and their directivity patterns.

Expressing the system transfer function as given above allows for a complete demarcation of the directional properties and positions of the source and microphone, wall absorption characteristics and room shape and size. This is especially useful in keeping computational cost low for cases where only a single aspect, such as source orientation, changes.

B. Stability

The stability of SDN follows from the fact that its recursive part, i.e. the backbone formed by the SDN nodes, is a fully connected DWN. The stability of lossless DWNs is, in turn, guaranteed by the fact that the network has a physical interpretation as a network of acoustic tubes [41], [9]. Indeed, the ideal physical system's total energy provides a Lyapunov function bounding the sum-of-squares in the numerical simulation (provided one uses rounding toward zero or error feedback in the computations) [41]. Furthermore, the network conserves its stability when losses due to wall absorption are included at the SDN nodes since the physical pressure (i.e. the sum of incoming and outgoing pressure waves) is always reduced relative to the lossless case.

C. Lossless Scattering Matrices

In this section, we explore possible choices for lossless scattering matrices and discuss their implications. First we present a complete parametrization of real lossless matrices.

The parameterization is an immediate corollary of the following theorem.

Theorem 1. *A real square matrix \mathbf{A} is diagonalizable if and only if it has the form*

$$\mathbf{A} = \mathbf{T}^{-1} \mathbf{\Lambda} \mathbf{T}, \quad (28)$$

where \mathbf{T} is a real invertible matrix, and $\mathbf{\Lambda}$ is a block diagonal matrix consisting of 1×1 blocks which are real eigenvalues of \mathbf{A} , and 2×2 blocks of the form

$$\begin{bmatrix} 0 & -r_i \\ r_i & 2r_i \cos(\theta_i) \end{bmatrix},$$

where $r_i e^{j\theta_i}$ are complex eigenvalues of \mathbf{A} which appear in pairs with their conjugates.

The theorem is proved in the appendix.

Corollary 2. *A real square matrix is lossless if and only if it has the form given by Theorem 1 with all eigenvalues on the unit circle.*

The corollary follows from the fact that a square matrix is lossless if and only if it is diagonalizable and all its values are on the unit circle [9].

Within the space of lossless matrices, a large degree of leeway is left for pursuing various physical or perceptual criteria. This can be achieved, for instance, by finding a lossless matrix \mathbf{A} which minimizes the distance from a real matrix \mathbf{D} that reflects some sought-after physical or perceptual properties. The design then amounts to constrained optimization:

$$\min_{\mathbf{A}} \|\mathbf{A} - \mathbf{D}\|_F^2, \quad (29)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, under the constraint that \mathbf{A} has the form given in Corollary 2. This most general case may, however, involve an excessive number parameters to optimise over. If we restrict the optimization domain to orthogonal matrices, solutions can be found without the need for numerical procedures. In particular, the following optimization result holds.

Theorem 3. *The solution to the following optimisation problem*

$$\operatorname{argmin}_{\mathbf{A}} \|\mathbf{A} - \mathbf{D}\|_F^2, \quad \mathbf{A}^T \mathbf{A} = \mathbf{I} \quad (30)$$

is given by $\mathbf{A} = \mathbf{U}\mathbf{V}^T$, where \mathbf{U} and \mathbf{V} are respectively the matrices of left and right singular vectors of \mathbf{D} .

A proof of this result can be found in [42].

Orthogonal matrices which are also circulant have the interpretation of performing all-pass circulant convolution of the incoming variables, which as discussed below, reduces computational complexity of the scattering operator. Furthermore, if a certain distribution of (unit-norm) eigenvalues is sought, the associated circulant matrix can be found by means of a single inverse fast Fourier transform (FFT) [9]. An in-depth study of such scattering matrices in the context of DWN reverberators has been presented in [9].

Householder reflection matrices, given by $\mathbf{A} = 2\mathbf{v}\mathbf{v}^T - \mathbf{I}$, $\|\mathbf{v}\| = 1$, are the subclass of orthogonal scattering matrices

commonly used in the context of DWN reverberators. As discussed in Section II-B, they enable a physical interpretation of the propagating variables as normalized pressure waves in a network of acoustic tubes. The optimization problem that minimizes the distance from a targeted scattering matrix in this case can be expressed as

$$\operatorname{argmin}_{\mathbf{v}} \|(2\mathbf{v}\mathbf{v}^T - \mathbf{I}) - \mathbf{D}\|_F^2, \quad \|\mathbf{v}\| = 1, \quad (31)$$

the solution of which is given by the following theorem.

Theorem 4. *The solution to the optimization problem in (31) is the singular vector corresponding to the largest singular value of matrix $\mathbf{D} + \mathbf{D}^T$.*

The theorem is proved in the appendix.

The isotropic scattering matrix, i.e. the particular case of a Householder reflection obtained for $\mathbf{v} = \mathbf{1}/\sqrt{K}$, can be physically interpreted as the scattering matrix which takes a reflection from one node and distributes its energy equally among all other nodes. This is the only orthogonal matrix which has this property, as stated by the following theorem.

Theorem 5. *If \mathbf{A} is an orthogonal matrix which scatters the energy from each incoming direction uniformly among all other directions, then it must have the form $\mathbf{A} = \pm \frac{2}{K} \mathbf{1}\mathbf{1}^T - \mathbf{I}$.*

The theorem is proved in the appendix. The isotropic matrix thus satisfies some optimality criteria and, as discussed below, allows for fast implementation. We will see in Section V-B that its special structure leads to a fast build up of echo density which is a perceptually desirable quality [43].

These different choices of scattering matrices have their implications on computational complexity. The computational complexity of the matrix-vector multiplication using general lossless and orthogonal matrices is $O[K^2]$ operations. Circulant lossless matrices require only $O[K \log(K)]$ operations [9]. The computational complexity associated with matrices which have the form of Householder reflections is further reduced to $O[K]$. Among general Householder reflections, which require $2K - 1$ additions and $2K$ multiplications, the isotropic scattering matrix only requires $2K - 1$ additions and 1 multiplication. Among lossless matrices, the ones that require the least operations are permutation matrices. However, we will see in Section V-B that permutation matrices lead to an insufficient echo density.

D. Interactivity and multichannel auralization

Interactive operation of the SDN reverberator is accomplished by updating the model to reflect changes in the positions and rotations of the source and microphone. This requires readjusting the positions of the wall nodes, and updating the delay line lengths and gains accordingly. It was observed in informal listening tests that updating the delay line lengths did not cause audible artifacts as long as microphone and source speeds are within reasonable limits.

The proposed model allows approximate emulation of coincident and non-coincident recording formats. Coincident formats (e.g. Ambisonics [18], higher-order ambisonics

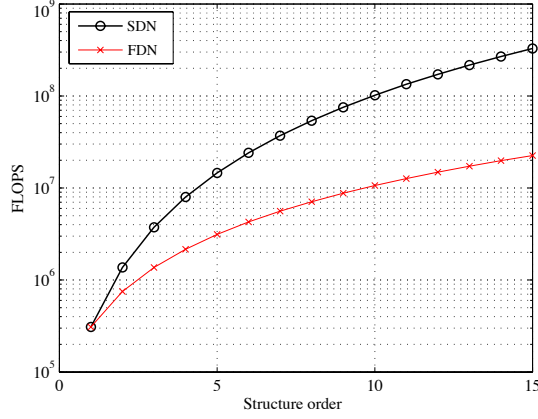


Fig. 10. Comparison of computational complexity for SDN and FDN as a function of the structure order, i.e. the size of the feedback matrix size for FDN and the number of neighboring nodes K for SDN (notice that the case of cuboid rooms corresponds to $K = 5$). The sampling frequency is $F_s = 44100$ Hz.

(HOA) [44], vector-base amplitude panning (VBAP) [45]) can be easily employed by adjusting the microphone gains $\Gamma_{M,k}(\theta)$ appropriately.

Non-coincident formats (e.g. [46], [47]) can be emulated by considering a separate SDN for each microphone. This results in a higher inter-channel decorrelation, which is what would actually occur in real recordings. However, the overall computational load also increases. If simulation speed is critical, an alternative approach would be to share the same set of wall nodes among all the SDNs, while creating dedicated node-to-microphone connection lines for each microphone.

E. Computational load

This section presents an analysis of the computational complexity of the proposed model in comparison to the two conceptually closest technologies, FDN and ISM. We use the number of floating point operations per second as an approximate indicator of the overall computational complexity. Furthermore, to simplify calculations we make the assumption that additions and multiplications carry the same cost. This approximation is motivated by the progressive convergence of computation time of various operations in modern mathematical processing units.

The number of floating point operations per second (FLOPS) performed by an SDN can be calculated as $F_s [2K^3 + (P+2)K^2 + K + 1]$, where P is the number of operations required by each wall filter for each sample.

Consider now an FDN with a $Q \times Q$ feedback matrix. From inspection of Fig. 1 the computational complexity can be shown to be $F_s [2Q^2 + (P+3)Q + 1]$ FLOPS.

Fig. 10 shows a comparison of the computational complexity of SDNs and FDNs. The sampling frequency is set to $F_s = 44100$ Hz and the filtering step consists of a simple frequency independent gain ($P = 1$) for both SDN and FDN. It may be observed that for the typical case of a 3D rectangular room ($K = 5$), SDN has around the same computational complexity of an FDN with a 12×12 feedback matrix.

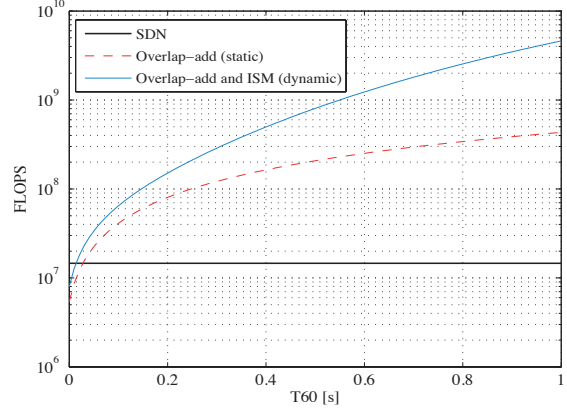


Fig. 11. Computational complexity of SDN in comparison to overlap-add convolution in both static and dynamic modes as a function of reverberation time. The static case represents the cost of the overlap-add convolution with a fixed, precomputed RIR. The dynamic case also includes the cost of calculating a new RIR and its FFT for each time frame. The sampling frequency is $F_s = 44100$ Hz.

Consider now the computational cost of the ISM method in its original time-domain implementation [13]. In ISM, each image source requires 10 floating-point operations¹ to calculate the time index and $1/r$ attenuation of the image, and 15 floating-point operations to calculate its attenuation due to wall absorption. This amounts to 25 floating-point operations in total for each image source. The number of image sources contained in an impulse response of length T_{60} seconds is approximately equal to the number of room cuboids that fit in a sphere with radius cT_{60} meters. This gives a computational complexity of around

$$25 \left\lceil \frac{\frac{4}{3}\pi(T_{60}c)^3}{L_x L_y L_z} \right\rceil \text{ FLOPS}, \quad (32)$$

where L_x , L_y and L_z are the room dimensions. Notice that the complexity is inversely proportional to the room volume—for a given reverberation time, smaller rooms have a higher computational cost.

The room impulse response generated by the ISM has to be convolved with the input signal to obtain the reverberated signal. In the case of real-time applications, this can be done efficiently using the overlap-add method [48]. The method calculates 2 FFTs of length N , N complex multiplications, 1 inverse FFT, and $\lceil T_{60}F_s \rceil$ real additions for each time frame. In order for the circular convolution to be equal to the linear convolution, N must satisfy $N \geq \lceil F_s/F_r \rceil + \lceil T_{60}F_s \rceil - 1$, where F_r is the frame refresh rate. In the best-case scenario where N is a power of 2, the asymptotic cost of each FFT is $6N \log_2 N$ [48]. Furthermore, each complex multiplication requires 4 real multiplications and 2 additions. Overall, the computational complexity of the overlap-add method is $F_r(18N \log_2 N + 6N + \lceil T_{60}F_s \rceil - 1)$ FLOPS. In the static case, the FFT of the impulse response can be precomputed, and the cost reduces to $F_r(12N \log_2 N + 6N + \lceil T_{60}F_s \rceil - 1)$ FLOPS.

¹Here we assume that exponentiations and square roots count as a single floating-point operation.

Fig. 11 compares the cost of SDN with the static and dynamic ISM. In the static case, both microphone and source are not moving. In the dynamic case, on the other hand, microphone and/or source are moving, and the ISM is run for each frame. The refresh rate is chosen such that the buffering delay is shorter than the maximum latency of a half-frame delay between the video and audio, as recommended by the ITU Recommendation BR.265-9 [49]. For a video running at 25 frames per second, this criterion gives a refresh rate of $F_r = 50$ Hz. The room size is the same as used in Allen and Berkley's paper of $10' \times 15' \times 12.5'$ feet [13]. It may be observed that for typical medium-sized rooms, SDN is from about 10 to 100 times faster than dynamic ISM. SDN is also significantly faster than (overlap-add) convolution alone.

F. Memory requirement

The required memory is determined by the number of taps of the delay lines. An upper bound for memory requirement Q can be easily found by observing that the length of each delay line is smaller than or equal to the distance between the two farthest points of the simulated space, giving:

$$Q \leq (N(N-1) + 2N + 1) \frac{qF_s}{c} R \text{ bits}, \quad (33)$$

where q is the number of bits per sample, and R is the maximum distance between any two points in the simulated space. The value of R in the case of a rectangular room is $R = \sqrt{L_x^2 + L_y^2 + L_z^2}$. For the more general case, R is the diameter of the bounding sphere of the room shape.

Observe in (33) that Q scales linearly with the room size. For a cubic room with a 5 m edge, $F_s = 40$ KHz, and $q = 32$ bytes per sample, the memory requirement is less than 170 KBytes, which is negligible for virtually every state-of-the-art platform.

V. ASSESSMENT

This section presents the results of assessments of SDNs in terms of perceptually based objective criteria.

As discussed in previous sections, first-order reflections are rendered correctly both in timing and amplitude by construction. Another cue, important for the perception of room volume and materials, is the reverberation time [37]. Section V-A presents an evaluation of the reverberation time both in frequency-independent and frequency-dependent cases. Section V-B focuses on the time evolution of echo density, which is related to the perceived time-domain texture of reverberation [43].

Here, ISM is used as a reference since it is the closest method among physical room models. More specifically, we use a C++ version of Allen and Berkley's original time-domain implementation [13].

A. Reverberation time

The parameter most commonly used to quantify the length of reverberation is T_{60} , which is defined as the time it takes for the room response to decay to 60 dB below its starting level. In this section, the T_{60} of the SDN network is compared with

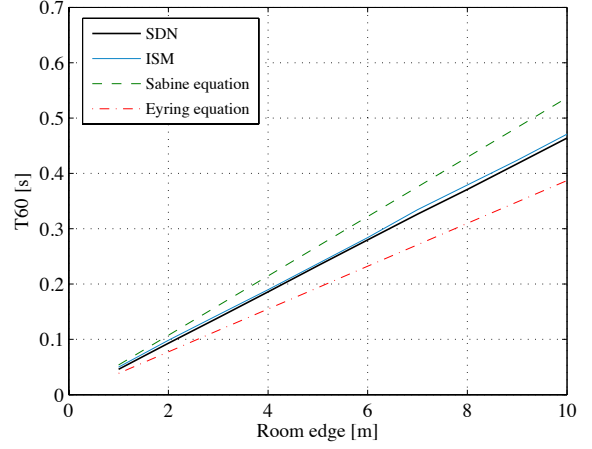


Fig. 12. Values of reverberation time T_{60} as a function of the edge of a cubic room. The wall absorption is $\alpha = 0.5$, and both microphone and source are positioned in the volumetric center of the room.

two well-known empirical formulas proposed by Sabine [50] and Eyring [51]:

$$T_{60,Sab} = \frac{0.161V}{\sum_i A_i \alpha_i}, \quad (34)$$

$$T_{60,Eyr} = -\frac{0.161V}{(\sum_i A_i) \log_{10}(1 - \sum_i A_i \alpha_i / \sum_i A_i)}, \quad (35)$$

where V is the room volume, A_i and α_i are the area and absorption coefficient of the i -th wall, respectively.

1) *Frequency-independent wall absorption*: Cubic rooms with different volumes and uniform frequency-independent absorption are simulated. In order to maintain the experimental conditions across different room sizes, both the source and the microphone are placed at the volumetric center of the room. Furthermore, the line-of-sight component was removed, as suggested in the ISO 3382 standard [52] for measuring the reverberation time in small enclosures. In Fig. 12, the reverberation time is shown as a function of the edge length for a room absorption coefficient $\alpha = 0.5$. It may be observed that SDN generates room impulse-responses which have reverberation times that increase linearly with the edge length. This is due to the larger average distance between the nodes, which in turn increases the mean free path of the structure.

The T_{60} values corresponding to the SDN reverberator are between the predictions given by Sabine and Eyring's formulas and are nearly identical to the ones produced by the ISM. The latter result may seem surprising if one considers that SDN, as opposed to the ISM, does not include attenuation due to spherical spreading (except for the initial first-order reflections and microphone taps). This apparent inconsistency is resolved intuitively by observing that spherical spreading is a lossless phenomenon: In ISM, the quadratic energy decrease ($1/r^2$) is compensated by the quadratic increase of the number of contributing image sources over time, and similarly to that, in DWN and SDN, "plane waves" are scattered losslessly at each node, thus conserving the energy.

In Fig. 13, the reverberation time is shown as a function of the absorption coefficient α . The enclosure was taken as a

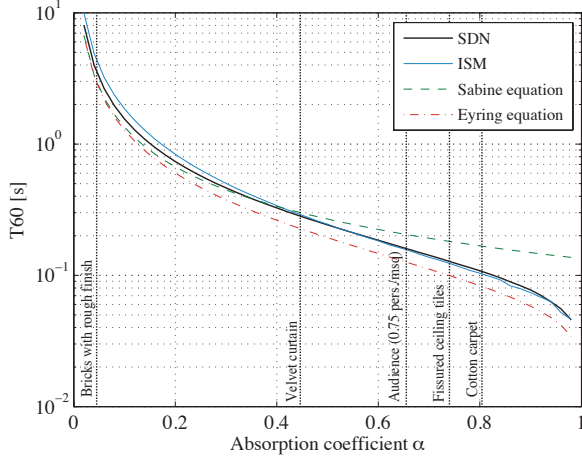


Fig. 13. Reverberation time values as a function of 98 values of the absorption coefficient α , equally spaced between 0.01 and 0.98 for SDN, ISM, and Sabine and Eyring predictions. The simulated enclosure is a cube with 4 m edge. T_{60} values are averaged across 10 source and microphone random positions. The 1 kHz absorption coefficient of various materials as measured by Vorländer in [38] are reported on top of the plot as reference.

cubic room with a 4 m edge, and results were averaged across 10 pairs of source-microphone positions. The coordinates were taken from a uniform random distribution and satisfied both requirements set in [52]: the microphone was at least 1 m away from the nearest wall, and the distance between the source and microphone was at least

$$d_{min} = 2\sqrt{\frac{V}{cT_{est}}}, \quad (36)$$

where T_{est} is a coarse estimation of the reverberation time. In these simulations, T_{est} was set using Sabine's formula.

It may be observed that for absorption coefficients higher than around $\alpha = 0.4$, SDN and ISM are nearly identical and are both between Sabine and Eyring's formulas. For high absorption coefficients, both SDN and ISM approach Eyring's formula, which is known to give more accurate predictions in that region [51]. For low absorption coefficients, SDN and ISM produce reverberation times longer than Sabine and Eyring's formulas, with SDN being closer to both.

2) *Frequency-dependent wall absorption*: Fig. 14 shows the result of a simulation where all walls mimic the frequency-dependent absorption of cotton carpet. The filters $H_i(z)$ are all set to be equal to a filter $H(z)$ which was implemented as a minimum-phase IIR filter with coefficients optimized by a damped Gauss-Newton method to fit the absorption coefficients reported by Vorländer in [38]. This procedure gave

$$H(z) = \frac{0.6876 - 1.9207z^{-1} + 1.7899 + z^{-2} - 0.5567z^{-3}}{1 - 2.7618z^{-1} + 2.5368z^{-2} - 0.7749z^{-3}}.$$

The source and microphone were positioned on the diagonal of a cubic room with 5 m edge. More specifically, they were positioned on the diagonal at a distance of $d_{min} = 2.96$ m from the center, as specified by equation (36).

In Fig. 14 the wall filter response is plotted together with the corresponding Sabine predictions in (34), which for the

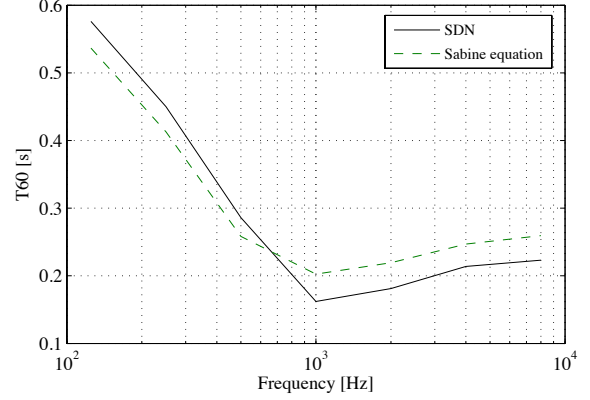


Fig. 14. Comparison of reverberation time in different octave bands for SDN and Sabine's formula prediction.

given room becomes

$$T_{60,Sab} = \frac{0.161V}{\sum_i A_i \alpha_i(\omega)} = \frac{0.161 \cdot 5}{6\alpha(\omega)} = \frac{0.161 \cdot 5}{6(1 - |H(e^{j\omega})|^2)}. \quad (37)$$

The simulated RIR was fed into an octave-band filter bank, and T_{60} values were calculated for each octave-band. As shown in Fig. 14, these measured T_{60} are very close to Sabine's formula prediction, thus confirming that the proposed model allows controlling the absorption behavior of wall materials explicitly.

Note that if explicit control of the reverberation time is sought, the prediction functions (34) or (35) can be inverted to obtain needed $\alpha(\omega)$. To this end, Eyring's formula (35) is preferable, since (34) may yield non-physical values for the absorption coefficient (i.e. $\alpha > 1$) of some acoustically "dead" rooms [51].

B. Echo density

The time evolution of echo density is commonly thought to influence the perceived time-domain *texture* of reverberation [43]. In an effort to quantify this perceptual attribute, Abel and Huang defined the normalized echo density (NED), which was found to have a very strong correlation with results of formal listening tests [43]. The NED is defined as the percentage of samples of room impulse response lying more than a standard deviation away from the mean in a given time window compared to that expected for Gaussian noise. A NED equal to 1 means that, within the considered window, the number of samples lying more than one standard deviation away from the mean is equal to the one observed with Gaussian noise.

Fig. 15 shows the time evolution of the NED obtained with the ISM and with the proposed model using three different scattering matrices. The scattering matrices are (a) the isotropic matrix, (b) a random orthogonal matrix, and (c) a random permutation. The random orthogonal matrix was obtained by setting the angles of a Givens-rotation parametrization of orthogonal matrices [53] at random. The simulated enclosure was a rectangular room with dimensions $3.2 \times 4.0 \times 2.7$ m, and results were averaged across 50 random pairs of source and

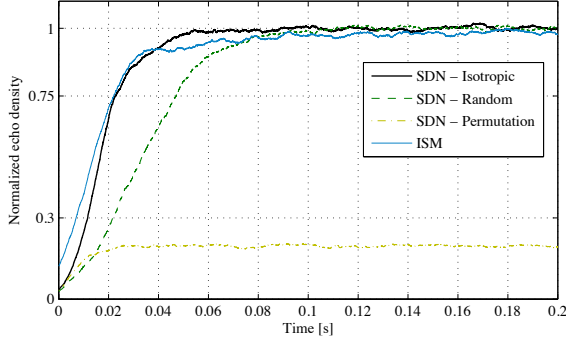


Fig. 15. Time evolution of the normalized echo density for ISM and SDN with various scattering matrices \mathbf{S} .

microphone positions. The wall gains were set to $\beta = -\sqrt{0.9}$ (wall absorption of $\alpha = 0.1$), with the negative sign being chosen in order to obtain a zero-mean reverberation tail with ISM.

Fig. 15 shows that the build-up of echo density of SDN is very close to that of ISM when the isotropic scattering matrix is employed. In particular, the NED values of 0.3 and 0.75, which were previously identified as breakpoints dividing three perceptually distinct groups [43], are reached at around the same delays by the two methods. Notice how the permutation matrix fails to reach a Gaussian-like reverberation. The random orthogonal matrix, on the other hand, does reach a Gaussian-like reverberation, but it takes longer to achieve the desired reverberation quality characterized by the 0.75 breakpoint.

VI. CONCLUSIONS

This paper presented a scalable and interactive reverberator design methodology termed Scattering Delay Network (SDN) that is capable of simulating the acoustics of a real room to a perceptually satisfying degree. The room is modeled by scattering nodes interconnected by bidirectional delay lines. These scattering nodes are positioned at the points where first-order reflections originate. In this way, the first-order reflections are simulated correctly, while a rich but less accurate reverberation tail is obtained. Examples were given which demonstrated that according to known perceptually based objective criteria, the proposed reverberator achieves the same reverberation quality as the image-source method while requiring one to two orders of magnitude lower computational load. Informal listening tests showed that the SDN reverberator provided natural sounding reverberation without artifacts.

APPENDIX

Towards proving Theorem 1, we first establish the following lemma.

Lemma 6. *Two diagonalizable matrices have the same eigenvalues if and only if they are similar.*

Proof of Lemma 6: The sufficiency is a well known result on similar matrices [54]. To prove the necessity, let us consider two diagonalizable matrices \mathbf{A} and \mathbf{B} which have

the same eigenvalues. Since \mathbf{A} is diagonalizable, it follows that $\mathbf{V}_A^{-1}\mathbf{A}\mathbf{V}_A = \mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is a diagonal matrix and \mathbf{V}_A is an invertible matrix. Hence, the following equality holds: $\mathbf{A}\mathbf{V}_A = \mathbf{V}_A\mathbf{\Lambda}$, which implies that $\mathbf{\Lambda}$ has eigenvalues of \mathbf{A} on its diagonal. Since \mathbf{B} is also diagonalizable and has the same eigenvalues as \mathbf{A} , it satisfies $\mathbf{V}_B^{-1}\mathbf{B}\mathbf{V}_B = \mathbf{\Lambda}$. Thus, $\mathbf{V}_A^{-1}\mathbf{A}\mathbf{V}_A = \mathbf{\Lambda} = \mathbf{V}_B^{-1}\mathbf{B}\mathbf{V}_B$, which further implies that $\mathbf{A} = (\mathbf{V}_B\mathbf{V}_A^{-1})^{-1}\mathbf{B}(\mathbf{V}_B\mathbf{V}_A^{-1})$, i.e. the two matrices are similar. ■

Using Lemma 6 we can now prove Theorem 1:

Proof of Theorem 1: First, observe that $\mathbf{\Lambda}$ is block diagonal, and that eigenvalues of each block are mutually distinct. Hence, each block of $\mathbf{\Lambda}$ is diagonalizable, and therefore $\mathbf{\Lambda}$ is itself diagonalizable [54] (over the field of complex numbers, \mathbb{C}). Thus, $\mathbf{\Lambda}$ has the same eigenvalues as \mathbf{A} and is diagonalizable. Since $\mathbf{\Lambda}$ and \mathbf{A} are both diagonalizable, and have the same eigenvalues, according to Lemma 6 they must be similar over \mathbb{C} . Moreover, since both $\mathbf{\Lambda}$ and \mathbf{A} are real, and similar over \mathbb{C} , they must be also similar over \mathbb{R} [54], that is, there must exist a real invertible matrix \mathbf{T} such that $\mathbf{A} = \mathbf{T}^{-1}\mathbf{\Lambda}\mathbf{T}$. On the other hand if \mathbf{A} has the form given in (28), it is diagonalizable since $\mathbf{\Lambda}$ is diagonalizable. ■

Proof of Theorem 4: Substituting the definition of a Householder transformation into the cost function in (31), here termed $\Phi(\mathbf{v})$, leads with simple algebraic manipulations to $\Phi(\mathbf{v}) = (\mathbf{v}^T\mathbf{v} - 1)\mathbf{v}^T\mathbf{v} - \mathbf{v}^T\mathbf{D}\mathbf{v} + \text{const}$. Minimizing $\Phi(\mathbf{v})$ subject to $\mathbf{v}^T\mathbf{v} = 1$ is therefore equivalent to maximising $\Phi_1(\mathbf{v}) = \mathbf{v}^T\mathbf{D}\mathbf{v}$ under the same constraint. The new cost function can be further expressed as

$$\Phi_1(\mathbf{v}) = \frac{1}{2}(\mathbf{v}^T\mathbf{D}\mathbf{v} + \mathbf{v}^T\mathbf{D}^T\mathbf{v}) = \frac{1}{2}\mathbf{v}^T(\mathbf{D} + \mathbf{D}^T)\mathbf{v}$$

and the unit norm vector which maximise it, is therefore the singular vector which corresponds to the largest singular value of $\mathbf{D} + \mathbf{D}^T$. ■

Proof of Theorem 5: The property that \mathbf{S} scatters energy from each node equally among all other nodes requires that all off-diagonal elements are identical, and thus that \mathbf{A} has the form

$$\mathbf{A} = \begin{bmatrix} a_1 & a_0 & \cdots & a_0 \\ a_0 & a_2 & \cdots & a_0 \\ \cdots & \cdots & \cdots & \cdots \\ a_0 & a_0 & \cdots & a_K \end{bmatrix}. \quad (38)$$

Orthogonality of \mathbf{A} requires that a_0, \dots, a_K satisfy

$$\begin{cases} a_i^2 + (K-1)a_0^2 = 1 & i = 1, \dots, K \\ a_0(a_i + a_j) + (K-2)a_0^2 = 0 & \forall i \neq j \end{cases}. \quad (39)$$

The first constraint can be written as $a_i = \pm\sqrt{1 - (K-1)a_0^2}$, which implies that all diagonal elements are identical in magnitude. The second constraint implies that if $a_0 \neq 0$ (the solution $a_0 = 0$ is ignored since it does not scatter energy), the diagonal elements have also the same sign. Solving (39) with $a_1 = \dots = a_K$ yields $a_1 = \pm(2-K)/K$ and $a_0 = \pm 2/K$, thus proving the theorem. ■

ACKNOWLEDGMENTS

The authors would like to thank Jonathan Abel and Geoffrey Robinson for their help and support.

REFERENCES

- [1] E. De Sena, H. Hacıhabiboğlu, and Z. Cvetković, "Scattering delay network: an interactive reverberator for computer games," in *Proc. 41st Audio Eng. Soc. Int. Conf.: Audio for Games*, London, UK, Feb. 2011.
- [2] H. Hacıhabiboğlu, E. De Sena, and Z. Cvetković, "Frequency-domain scattering delay networks for simulating room acoustics in virtual environments," in *Signal-Image Technology and Internet-Based Systems (SITIS)*, 2011, 2011, pp. 180–187.
- [3] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Trans. on Audio, Speech and Language Process.*, vol. 20, no. 3, pp. 1421–1448, Jul. 2012.
- [4] M. R. Schroeder and B. F. Logan, "Colorless artificial reverberation," *IRE Trans. on Audio*, vol. AU-9, pp. 209–214, Nov. 1961.
- [5] M. R. Schroeder, "Natural sounding artificial reverberation," *J. Audio Eng. Soc.*, vol. 10, no. 3, pp. 219–233, 1962.
- [6] M. Gerzon, "Unitary (energy-preserving) multichannel networks with feedback," *IET Electronics Letters*, vol. 12, no. 11, pp. 278–279, 1976.
- [7] J. Stautner and M. Puckette, "Designing multichannel reverberators," *Computer Music J.*, vol. 6, no. 1, pp. 52–65, 1982.
- [8] J.-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," presented at the 104th Audio Eng. Soc. Conv., Preprint #3030, Paris, France, Feb. 1991.
- [9] D. Rocchesso and J. O. Smith, "Circulant and elliptic feedback delay networks for artificial reverberation," *IEEE Trans. on Speech and Audio Process.*, vol. 5, no. 1, pp. 51–63, January 1997.
- [10] J. O. Smith, "A new approach to digital reverberation using closed waveguide networks," in *Proc. 11th Int. Comput. Music Conf.*, Burnaby, Canada, 1985, pp. 47–53.
- [11] M. Vorlander, *Auralization - Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer-Verlag, 2008.
- [12] L. Savioja, J. Backman, A. Järvinen, and T. Takala, "Waveguide mesh method for low-frequency simulation of room acoustics," *ica95*, pp. 637–640, Jun. 1995.
- [13] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.
- [14] S. Bilbao and J. O. Smith, III, "Finite difference schemes and digital waveguide networks for the wave equation: Stability, passivity, and numerical dispersion," *IEEE Trans. on Speech and Audio Processing*, vol. 11, no. 3, pp. 255–266, May 2003.
- [15] D. Murphy, A. Kelloniemi, J. Mullen, and S. Shelley, "Acoustic modeling using the digital waveguide mesh," *Signal Processing Magazine, IEEE*, vol. 24, no. 2, pp. 55–66, 2007.
- [16] M. Karjalainen and C. Erku, "Digital waveguides versus finite difference structures: Equivalence and mixed modeling," *EURASIP Journal on Applied Signal Processing*, vol. 2004, pp. 978–989, 2004.
- [17] M. Zyda, "From visual simulation to virtual reality to games," *Computer*, vol. 38, no. 9, pp. 25–32, September 2005.
- [18] M. A. Gerzon, "Ambisonics in multichannel broadcasting and video," *J. Audio Eng. Soc.*, vol. 33, no. 11, pp. 859–871, November 1985.
- [19] M. M. Boone, U. Horbach, and W. P. J. Bruijn, "Spatial sound-field reproduction by wave-field synthesis," *J. Audio Eng. Soc.*, vol. 43, no. 12, pp. 1003–1012, December 1995.
- [20] H. Möller, "Fundamentals of binaural technology," *Appl. Acoust.*, vol. 36, no. 3–4, pp. 171–218, 1991.
- [21] *SMPTE 2036-2-2008: Ultra High Definition Television – Audio Characteristics and Audio Channel Mapping for Program Production*. Society of Motion Picture and Television Engineers, 2008.
- [22] D. Rocchesso, "The ball within the box: a sound-processing metaphor," *Computer Music Journal*, pp. 47–57, 1995.
- [23] P. Huang, M. Karjalainen, and J. O. Smith, "Digital Waveguide Networks for Room Response Modeling and Synthesis," presented at the 118th Audio Eng. Soc. Conv., Preprint # 6394, Barcelona, Spain, May 2005.
- [24] J. Borish, "Extension of the image model to arbitrary polyhedra," vol. 75, no. 6, pp. 1827–1836, 1984.
- [25] C. Faller and F. Menzer, "Unitary Matrix Design for Diffuse Jot Reverberators," in *Proc. 128th Conv. Audio Eng. Soc.*, Preprint # 7984, London, UK, 2010.
- [26] J. O. Smith, *Physical audio signal processing*. W3K Publishing, 2010, URL: <https://ccrma.stanford.edu/~jos/pasp/>.
- [27] G. Evangelista and F. Eckerholm, "Player-instrument interaction models for digital waveguide synthesis of guitar: Touch and collisions," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 4, pp. 822–832, 2010.
- [28] J. Sheaffer, M. van Walstijn, and B. Fazenda, "Physical and numerical constraints in source modeling for finite difference simulation of room acoustics," *The Journal of the Acoustical Society of America*, vol. 135, no. 1, pp. 251–261, 2014.
- [29] D. T. Murphy, A. Southern, and L. Savioja, "Source excitation strategies for obtaining impulse responses in finite difference time domain room acoustics simulation," *Applied Acoustics*, vol. 82, pp. 6–14, 2014.
- [30] D. T. Murphy, "Digital waveguide mesh topologies in room acoustics modelling," Ph.D. dissertation, University of York, 2000.
- [31] S. D. Bilbao, *Wave and scattering methods for numerical simulation*. Wiley, 2004.
- [32] A. Kelloniemi, V. Välimäki, and L. Savioja, "Simulation of room acoustics using 2-D digital waveguide meshes," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP-06)*, vol. 5, Toulouse, France, May 2006.
- [33] H. Hacıhabiboğlu, B. Günel, and Z. Cvetković, "Simulation of directional microphones in digital waveguide mesh-based models of room acoustics," *IEEE Trans. on Audio, Speech and Language Process.*, vol. 18, no. 2, pp. 213–223, February 2010.
- [34] J. C. Strikwerda, *Finite Difference Schemes and Partial Differential Equations*, 2nd ed. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2004.
- [35] L. Savioja and V. Välimäki, "Interpolated rectangular 3-D digital waveguide mesh algorithms with frequency warping," *IEEE Trans. on Speech and Audio Process.*, vol. 11, no. 6, pp. 783–790, November 2004.
- [36] H. Kuttruff, *Room Acoustics*, 4th ed. SPON Press, 2000.
- [37] F. Rumsey, *Spatial audio*. Focal Press, 2001.
- [38] M. Vorlander, *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer Verlag, 2008.
- [39] J. Huopaniemi, L. Savioja, and M. Karjalainen, "Modeling of reflections and air absorption in acoustical spaces a digital filter design approach," in *Proc. IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA'97)*, New Paltz, NY, USA, October 1997.
- [40] J. A. Moorer, "About this reverberation business," *Computer Music Journal*, vol. 3, no. 2, pp. 13–28, 1979.
- [41] J. O. Smith and D. Rocchesso, "Aspects of digital waveguide networks for acoustic modeling applications," *Web published at <http://ccrma.stanford.edu/jos/wgj>*, 1997.
- [42] Z. Zhang, "A flexible new technique for camera calibration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [43] P. Huang and J. S. Abel, "Aspects of reverberation echo density," presented at the 123rd Audio Eng. Soc. Conv., Preprint #7163, New York, USA, Oct. 2007.
- [44] M. Poletti, "A unified theory of horizontal holographic sound systems," *J. Audio Eng. Soc.*, vol. 48, no. 12, pp. 1155–1182, December 2000.
- [45] V. Pulkki, "Virtual sound source positioning using vector-base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, June 1997.
- [46] J. D. Johnston and Y. H. Lam, "Perceptual soundfield reconstruction," presented at the 109th Audio Eng. Soc. Conv., Preprint #2399, Los Angeles, CA, USA, September 2000.
- [47] E. De Sena, H. Hacıhabiboğlu, and Z. Cvetković, "Analysis and design of multichannel systems for perceptual sound field reconstruction," *IEEE Trans. on Audio, Speech and Language Process.*, vol. 21, no. 8, pp. 1653–1665, August 2013.
- [48] A. Oppenheim, R. Schaffer, J. Buck *et al.*, *Discrete-time signal processing*. Prentice hall Englewood Cliffs, NJ, 1989, vol. 2.
- [49] *Recomm. BR.265-9, Operating practices for the international exchange of programmes** on film for television use*. International Telecommunications Union, 2004.
- [50] W. C. Sabine, "Reverberation," in *Collected papers on acoustics*. Harvard university press, 1922.
- [51] C. F. Eyring, "Reverberation time in "dead" rooms," *J. Acoust. Soc. Am.*, vol. 1, p. 168, 1930.
- [52] Various, *Standard 3382. Acoustics—Measurement of the reverberation time of rooms with reference to other acoustical parameters*, 1997.
- [53] M. Vetterli and J. Kovačević, *Wavelets and subband coding*. Prentice Hall PTR Englewood Cliffs, New Jersey, 1995, vol. 152.
- [54] L. Hogben, *Handbook of linear algebra*. CRC Press, 2006.